

# WantWords: An Open-source Online Reverse Dictionary System

Fanchao Qi<sup>1,2\*</sup>, Lei Zhang<sup>2\*†</sup>, Yanhui Yang<sup>2†</sup>, Zhiyuan Liu<sup>1,2,3</sup>, Maosong Sun<sup>1,2,3</sup>

<sup>1</sup>Department of Computer Science and Technology, Tsinghua University

<sup>2</sup>Institute for Artificial Intelligence, Tsinghua University

Beijing National Research Center for Information Science and Technology

<sup>3</sup>Beijing Academy of Artificial Intelligence

qfc17@mails.tsinghua.edu.cn, zhanglei9003@gmail.com

flutter0696@gmail.com, {liuzy, sms}@tsinghua.edu.cn

## Abstract

A reverse dictionary takes descriptions of words as input and outputs words semantically matching the input descriptions. Reverse dictionaries have great practical value such as solving the tip-of-the-tongue problem and helping new language learners. There have been some online reverse dictionary systems, but they support English reverse dictionary queries only and their performance is far from perfect. In this paper, we present a new open-source online reverse dictionary system named WantWords (<https://wantwords.thunlp.org/>). It not only significantly outperforms other reverse dictionary systems on English reverse dictionary performance, but also supports Chinese and English-Chinese as well as Chinese-English cross-lingual reverse dictionary queries for the first time. Moreover, it has user-friendly front-end design which can help users find the words they need quickly and easily. All the code and data are available at <https://github.com/thunlp/WantWords>.

## 1 Introduction

Opposite to a regular (forward) dictionary that provides definitions for query words, a *reverse dictionary* (Sierra, 2000) returns words semantically matching the query descriptions. In Figure 1, for example, a regular dictionary tells you the definition of “expressway” is “a wide road that allows traffic to travel fast”, while a reverse dictionary outputs “expressway” and other semantically similar words like “freeway” which match the query description “a road where cars go very quickly without stopping” you input.

Reverse dictionaries are useful in practical applications. First and foremost, they can effectively solve the *tip-of-the-tongue* problem (Brown and

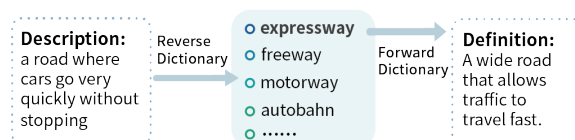


Figure 1: An example illustrating what a regular (forward) dictionary and a reverse dictionary are.

McNeill, 1966), namely the phenomenon of failing to retrieve a word from memory. Many people frequently suffer the problem, especially those who write a lot such as writers, researchers and students. With the help of reverse dictionaries, people can quickly and easily find the words that they need but temporarily forget.

In addition, reverse dictionaries are helpful to new language learners who grasp a limited number of words. They will know and learn some new words that have the meanings they want to express by using a reverse dictionary. Also, reverse dictionaries can help word selection (or word dictionary) anomia patients, people who can recognize and describe an object but fail to name it due to neurological disorder (Benson, 1979).

Currently, there are mainly two online reverse dictionaries, namely OneLook<sup>1</sup> and ReverseDictionary.<sup>2</sup> Their performance is far from perfect. Further, both of them are closed-source and only support English reverse dictionary queries.

To solve these problems, we design and develop a new online reverse dictionary system named WantWords, which is totally open-source. WantWords is mainly based on our proposed multi-channel reverse dictionary model (Zhang et al., 2020), which achieves state-of-the-art performance on an English benchmark dataset. Our system uses an improved version of the multi-channel reverse dictionary model and incorporates some

\*Indicates equal contribution

†Work done during internship at Tsinghua University

<sup>1</sup><https://onelook.com/thesaurus/>

<sup>2</sup><https://reversedictionary.org/>

engineering tricks to handle extreme cases. Evaluation results show that with these improvements, our system achieves higher performance. Besides, our system supports Chinese reverse dictionary queries and Chinese-English as well as English-Chinese cross-lingual reverse dictionary queries, all of which are realized for the first time. Finally, our system is very user-friendly. It includes multiple filters and sort methods, and can automatically cluster the candidate words, all of which help users find the target words as quickly as possible.

## 2 Related Work

There are mainly two methods for reverse dictionary building. The first one is based on sentence matching (Bilac et al., 2004; Zock and Bilac, 2004; Méndez et al., 2013; Shaw et al., 2013). Its main idea is to return the words whose dictionary definitions are most similar to the query description. Although effective in some cases, this method cannot cope with the problem that human-written query descriptions might differ widely from dictionary definitions.

The second method uses a neural language model (NLM) to encode the query description into a vector in the word embedding space, and returns the words with the closest embeddings to the vector of the query description (Hill et al., 2016; Morinaga and Yamaguchi, 2018; Kartsaklis et al., 2018; Hedderich et al., 2019; Pilehvar, 2019). Performance of this method depends largely on the quality of word embeddings. Unfortunately, according to Zipf’s law (Zipf, 1949), many words are low-frequency and usually have poor embeddings.

To tackle this issue of the NLM-based method, we proposed a multi-channel reverse dictionary model (Zhang et al., 2020). This model is composed of a sentence encoder, more specifically, a bi-directional LSTM (BiLSTM) (Hochreiter and Schmidhuber, 1997) with attention (Bahdanau et al., 2015), and four characteristic predictors. The four predictors are used to predict the part-of-speech, morphemes, word category and sememes<sup>3</sup> of the target word according to the query description, respectively. The incorporation of the characteristic predictors can help find the target words with poor embeddings and exclude wrong words with similar embeddings to the target words, such

<sup>3</sup>A sememe is defined as the minimum semantic units of human languages (Bloomfield, 1926). The meaning of a word can be expressed by several sememes.

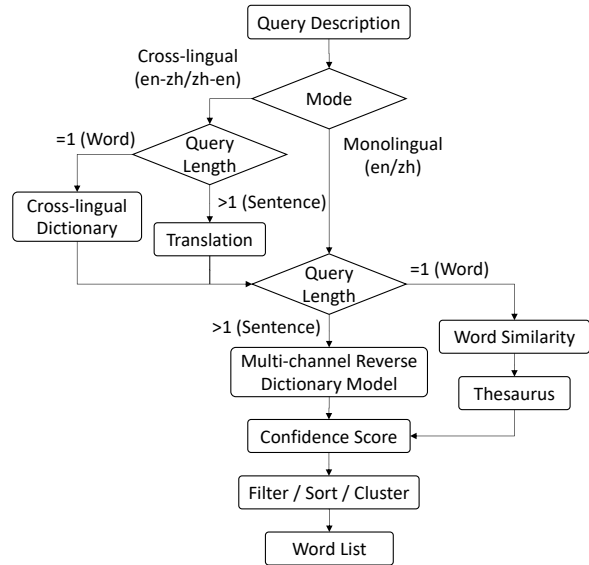


Figure 2: Workflow of WantWords.

as antonyms. Experimental results have demonstrated that our multi-channel reverse dictionary model achieves state-of-the-art performance. In WantWords, we employ an improved version of it that yields better results.

## 3 System Architecture

In this section, we describe the system architecture of WantWords. We first give an overview of its workflow, then we detail the improved multi-channel reverse dictionary model, and finally we introduce its front-end design.

### 3.1 Overall Workflow

The workflow of WantWords is illustrated in Figure 2. There are two reverse dictionary modes, namely monolingual and cross-lingual modes. In the monolingual mode, if the query description is longer than one word, it will be fed into the multi-channel reverse dictionary model directly, which calculates a confidence score for each candidate word in the vocabulary; if the query description is just a word, the confidence score of each candidate word is mostly based on the cosine similarity between the embeddings of the query word and candidate word.

In the cross-lingual mode, where the query descriptions are in the *source* language and the target words are in the *target* language, if the query description is longer than one word, it will be translated into the target language first and then processed in the monolingual mode of the target language; if the query description is just a word, cross-lingual dictionaries will be consulted for the target-

language definitions of the query word, and then the definitions are fed into the multi-channel reverse dictionary model to calculate candidate words’ confidence scores.

After obtaining confidence scores, all candidate words in the vocabulary will be sorted by descending confidence scores and listed as system output. The words in the query description are excluded since they are unlikely to be the target word. Different filters, other sort methods and clustering may be further employed to adjust the final results.

### 3.2 Multi-channel Reverse Dictionary Model

The multi-channel reverse dictionary model (MRDM) is the core module of our system. We use an improved version of MRDM that employs BERT (Devlin et al., 2019) rather than BiLSTM as the sentence encoder. Figure 3 illustrates the model.

For a given query description, MRDM calculates a confidence score for each candidate word in the vocabulary. The confidence score is composed of five parts:

(1) The first part is word score. To obtain it, the input query description is first encoded into a sentence vector by BERT, then the sentence vector is mapped into the space of word embeddings by a single-layer perceptron, and finally word score is the dot product of the mapped sentence vector and the candidate word’s embedding.

(2) The second part is part-of-speech (PoS) score, which is based on the prediction for the PoS of the target word. MRDM first calculates a prediction score for each PoS tag by feeding the sentence vector into a single-layer perceptron, and then a candidate word’s PoS score is the sum of the prediction scores of all its PoS tags.

(3) The third part is category score, which is related to the category of the target word and can be obtained in a similar way to PoS score.

(4) The fourth part is morpheme score, which is supposed to capture the morphemes of the target word. Each token of the input query description corresponds to a hidden state as the output of BERT. MRDM first feeds each hidden state into a single-layer perceptron to obtain a local morpheme prediction score, then does max-pooling over all the local morpheme prediction scores to obtain a prediction score for each morpheme, and finally a candidate word’s morpheme score is the sum of the prediction scores of all its morphemes.

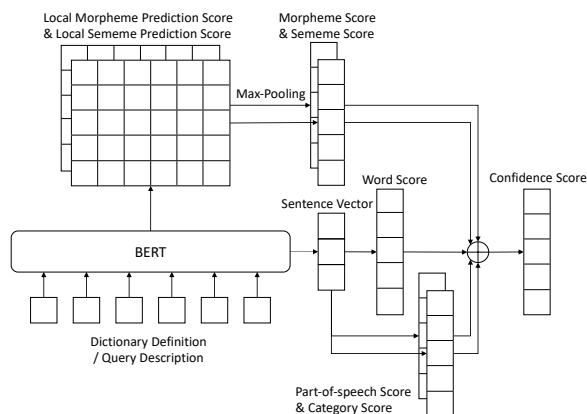


Figure 3: Revised version of the multi-channel reverse dictionary model.

(5) The fifth part is sememe score, which is based on the prediction for the sememes of the target word. Sememe score can be calculated in a similar way to morpheme score.

We use the official pre-trained BERT models for both English and Chinese.<sup>4</sup> As for fine-tuning (training) for English, we use the dictionary definition dataset created by Hill et al. (2016), which contains about 100,000 words and 900,000 word-definition pairs extracted from five dictionaries. For fine-tuning (training) for Chinese, we build a large-scale dictionary definition dataset based on the dataset created by Zhang et al. (2020). It contains 137,174 words and 270,549 word-definition pairs, where the definitions are extracted from several authoritative Chinese dictionaries including *Modern Chinese Dictionary*, *Xinhua Dictionary* and *Chinese Idiom Dictionary* as well as an open-source dictionary dataset.<sup>5</sup>

MRDM requires some other resources, and we simply follow the settings in Zhang et al. (2020). Specifically, for English, we use Morfessor (Virpioja et al., 2013) to segment words into morphemes, WordNet (Miller, 1995) to obtain PoS and word category information, and OpenHowNet<sup>6</sup> (Qi et al., 2019) to obtain sememe information. As for Chinese, we simply use Chinese characters as morphemes. We utilize the PoS tags in *Modern Chinese Dictionary*. In addition, we use HIT-IR Tongyici Cilin<sup>7</sup> and OpenHowNet to obtain word category and sememe information, respectively.

<sup>4</sup><https://github.com/google-research/bert>

<sup>5</sup><https://github.com/pwxcoo/chinese-xinhua>

<sup>6</sup><https://github.com/thunlp/OpenHowNet>

<sup>7</sup>[https://github.com/yaleimeng/Final\\_word\\_Similarity/tree/master/cilin](https://github.com/yaleimeng/Final_word_Similarity/tree/master/cilin)

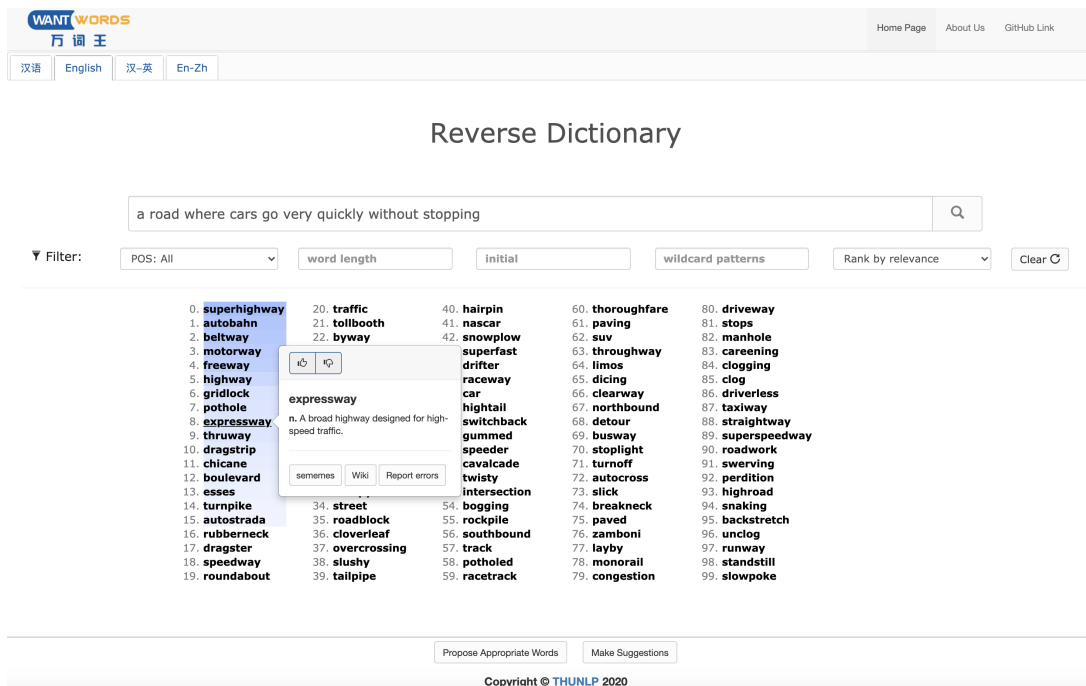


Figure 4: Front-end design of WantWords in the English monolingual mode.

### 3.3 One-word Query in the Monolingual Mode

In the monolingual reverse dictionary mode, in the case where the query description is a single word, we simply use word embedding similarity to calculate the confidence scores of candidate words in the vocabulary, rather than feed the query word into MRDM. We also take the synonyms into consideration and double the confidence score of a candidate word if it is a synonym of the query word. We use WordNet and HIT-IR Tongyici Cilin as English and Chinese thesauri, respectively.

### 3.4 The Cross-lingual Mode

In the cross-lingual mode, a query description longer than one word is first translated into the target language using Baidu Translation API<sup>8</sup>, and then the translated query description is processed in the same procedure as the monolingual mode.

As for a one-word query description, we do not utilize machine translation because existing translation APIs cannot return all the possible translation results, especially for polysemous query words, which may impair system performance. Instead, we consult cross-lingual dictionaries for definitions in the target language of the query word, and feed all the definitions into the target-language MRDM. Specifically, we use *StarDict* and *LangDao* English-Chinese Dictionaries in the

<sup>8</sup><https://fanyi-api.baidu.com/>

English-Chinese mode and *LangDao*, *CEDICT*, and *MDBG* Chinese-English dictionaries in the Chinese-English mode. We concatenate multiple dictionary definitions before feeding into MRDM.

### 3.5 Front-end Design

The front-end design of WantWords is simple and user-friendly, as shown in Figure 4. After inputting a query description in the textbox in the center of the system web page and clicking the “Search” button, one hundred candidate words will be listed in descending order of confidence scores. The words with confidence scores higher than a threshold have a background color whose shade is proportional to the confidence score.

A tool bar will appear below the textbox. Users can filter the candidate words by different filters in the tool bar. Specifically, for English candidate words, there are PoS, word length, initial and wildcard pattern filters; for Chinese candidate words, there are word length, total stroke number, wildcard pattern, pinyin initials, PoS and rhyme filters. These filters can help users find the word they need as quickly as possible.

In the tool bar, users can also change the sort method of the candidate words. Users can sort the English candidate words in regular or reverse alphabetical order and by word length, and Chinese candidate words in regular or reverse pinyin alphabetical order and by total stroke number. Besides,

WantWords supports dividing candidate words into six clusters, where we use k-means clustering algorithm in the word embedding space. The sort methods and clustering are also beneficial to quickly finding the target word.

Considering the cases where users, especially new language learners, do not know rather than forget a word, our system provides definitions for candidate words. Users can click a candidate word to invoke a floating window that displays the PoS and definition of the word. The displayed definitions of English and Chinese words are from WordNet and the open-source Chinese dictionary dataset respectively, both of which are freely available.

Finally, our system has quick feedback channels to collect real-world data. Due to the lack of human-written description data, existing reverse dictionary systems can only utilize dictionary definitions in training. However, dictionary definitions are usually different from human-written descriptions, which affects the performance of reverse dictionaries. Therefore, we design some feedback channels to collect users' feedback, aiming to use it to improve our system. Specifically, users can choose between "Matched Well" and "Not Matched" in the floating window of a candidate word to give their opinions about the candidate word. In addition, users can directly propose appropriate words matching the query description.

## 4 Evaluation

In this section, we evaluate the reverse dictionary performance of WantWords. We conduct both monolingual (English and Chinese) and cross-lingual (English-Chinese and Chinese-English) reverse dictionary evaluations.

### 4.1 Datasets

In the evaluation of English monolingual reverse dictionary performance, we use two test sets including (1) **Definition** set, which contains 500 pairs of words and WordNet definitions that are randomly selected and have been excluded from the training set; and (2) **Description** set, which comprises 200 pairs of words and human-written descriptions and is a benchmark dataset created by Hill et al. (2016).

As for Chinese, we use three test sets: (1) **Definition** set, which contains 2,000 pairs of words and dictionary definitions that are selected at random and do not exist in the training set; (2) **Description** set, which is composed of 200 word-description

pairs given by Chinese native speakers and is built by Zhang et al. (2020); and (3) **Question** set, which collects 272 real-world Chinese exam question-answers of writing the right word given a description from the Internet and is also created by Zhang et al. (2020).

To evaluate cross-lingual reverse dictionary performance, we build two test sets based on the two monolingual Description sets. We manually translate the word of each word-description pair in the English Description sets into Chinese to obtain the **English-Chinese Description** set, which is composed of 200 pairs of English descriptions and Chinese words. In a similar way, we construct the **Chinese-English Description** set, which contains 200 pairs of Chinese descriptions and English words.

### 4.2 Baseline Methods

We choose two existing online reverse dictionary systems, namely OneLook and ReverseDictionary, and two reverse dictionary models, namely original MRDM and BERT, as baseline methods.

OneLook and ReverseDictionary can only support English monolingual reverse dictionary queries. MRDM, as mentioned before, is the current state-of-the-art reverse dictionary model (Zhang et al., 2020) and mainly differs from WantWords in the sentence encoder (BiLSTM vs BERT) and engineering tricks (e.g., special processing for one-word queries) to handle one-word queries. As for BERT, it does not have extra characteristic predictors and engineering tricks as compared to WantWords. MRDM and BERT are trained with the same training sets as WantWords to respond English and Chinese reverse dictionary queries, respectively. They can also support cross-lingual reverse dictionary queries processed with the same procedure as the cross-lingual mode of WantWords.

### 4.3 Evaluation Metrics

Following previous work (Hill et al., 2016; Zhang et al., 2020), we use four evaluation metrics: the median rank of the target words in the final word lists (lower better) and the accuracy that the target words appear in top 1/10/100 (acc@1/10/100, higher better). Every experiment is run five times, and we report the average results. We also conduct Student's t-test to measure the significance of performance difference.

Model	En_Definition	En_Description	Zh_Definition	Zh_Description	Zh_Question
OneLook	– –	6 .33/.54/.76	– –	– –	– –
ReverseDictionary	– –	4 .30/.64/.80	– –	– –	– –
MRDM	53 .08/.29/.59	3 .31/.65/.88	8 .21/.51/.76	4 .27/.60/.85	1 .50/.79/.91
BERT	34 .09/.34/.61	<u>2</u> .33/.76/.93	13 .13/.45/.72	5 .23/.62/.86	1 .49/.79/.91
WantWords	<b>19</b> .10/.38/.72	<u>2</u> .36/.75/.92	<b>7</b> .22/.54/.77	<b>2</b> .37/.74/.91	<b>0</b> .60/.82/.93

Table 1: Evaluation results of English and Chinese monolingual reverse dictionaries (median rank and acc@1/10/100). The boldfaced results denote significant dominance, and the underlined results indicate insignificant difference, where the statistical significance threshold of p-value is 0.05. The same is true for Table 2.

Model	en-zh	zh-en
MRDM	40 .12/.31/.63	<u>8</u> .20/.52/.76
BERT	<b>16</b> .14/.40/.75	7 .21/.54/.76
WantWords	19 .14/.38/.76	<u>8</u> .22/.53/.78

Table 2: Evaluation results of cross-lingual reverse dictionaries (median rank and acc@1/10/100).

#### 4.4 Evaluation Results

The monolingual reverse dictionary evaluation results of WantWords and baseline methods are shown in Table 1. OneLook and ReverseDictionary have stored all the WordNet definitions, and we cannot exclude the word-definition pairs in the Definition set from their databases. Therefore, they can be evaluated on the Description set only.

We observe that WantWords basically performs better than all the baseline methods on all the five test sets. On the English benchmark test set Description, WantWords completely outperforms the two existing online systems and achieves new state-of-the-art performance. On the three Chinese test sets, WantWords also yields significantly better results than the two baseline methods.

Table 2 shows the cross-lingual reverse dictionary evaluation results of WantWords and two baseline methods. We find that the performance of three models is similar and much poorer than that on corresponding monolingual datasets. We conjecture that the unsatisfying translation quality seriously affects final performance, based on our observation that translations of some query descriptions are inaccurate and even ungrammatical.

#### 4.5 Case Study

Table 3 shows two English reverse dictionary cases, where the query descriptions and output word lists of three reverse dictionary systems are displayed. In the first case, WantWords finds 8 correct words among top 10 while the other two systems find none among top 15. In the second case, OneLook and WantWords use the PoS filter to retain verbs only. After filtering, the target word “receive” is

System	Results
Query: a road where cars go very quickly without stopping	
OneLook	station stationing stations pass passed stump stumped stumping stumps shoulder turnout flash run turn turned
ReverseDictionary	run garage scoot tipple rush direct shoulder flash tollbooth interchange resort hie drive station cessation
WantWords	<b>superhighway autobahn beltway motorway freeway highway</b> gridlock pothole <b>expressway thruway</b> dragstrip
Query: when somebody gives something to you and afterwards you have it	
OneLook	can lie form catch else greyhound retribution employ tire snap order find assure <b>employed face starter bolt</b> reach
ReverseDictionary	have satisfaction lose claim deposit guess license repent change ration misappropriate charge carry leave remember
WantWords	<b>gift pawning gifting loaning beneficiary payment buying giving giver how cash bonus refund purchase receive</b>

Table 3: Two English reverse dictionary cases. The boldfaced words are correct answers while the words struck through are filtered out by PoS filter (verb).

ranked top 3 in the word list of WantWords while OneLook still cannot find any correct words among top 6. ReverseDictionary has no filter and none of correct words appear among top 15.

## 5 Conclusion and Future work

In this paper, we present WantWords, an online reverse dictionary system, which achieves state-of-the-art performance on an English reverse dictionary benchmark dataset. Besides, it supports Chinese and English-Chinese as well as Chinese-English cross-lingual reverse dictionary queries for the first time. In the future, we will try to incorporate multi-word expressions and idioms in the system. Also, we will work on improving cross-lingual reverse dictionary performance by bilingual word embeddings or multilingual BERT.

## Acknowledgements

This work is supported by the Major Program of the National Social Science Fund of China (No. 18ZDA238), the National Key Research and Development Program of China (No. 2020AAA0105200 and 2018YFB1004503), the National Natural Science Foundation of China (NSFC, No. 61732008), the NExT++ project from the National Research Foundation, Prime Minister’s Office, Singapore under its IRC@Singapore Funding Initiative, and Beijing Academy of Artificial Intelligence (BAAI).

## References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of ICLR*.
- D Frank Benson. 1979. Neurologic correlates of anomia. In *Studies in Neurolinguistics*, pages 293–328. Elsevier.
- Slaven Bilac, Wataru Watanabe, Taiichi Hashimoto, Takenobu Tokunaga, and Hozumi Tanaka. 2004. Dictionary search based on the target word description. In *Proceedings of NLP*.
- Leonard Bloomfield. 1926. A set of postulates for the science of language. *Language*, 2(3):153–164.
- Roger Brown and David McNeill. 1966. The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5(4):325–337.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*.
- Michael A Hedderich, Andrew Yates, Dietrich Klakow, and Gerard de Melo. 2019. Using multi-sense vector embeddings for reverse dictionaries. In *Proceedings of IWCS*.
- Felix Hill, Kyunghyun Cho, Anna Korhonen, and Yoshua Bengio. 2016. Learning to understand phrases by embedding the dictionary. *TACL*, 4:17–30.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.
- Dimitri Kartsaklis, Mohammad Taher Pilehvar, and Nigel Collier. 2018. Mapping text to knowledge graph entities using multi-sense lstms. In *Proceedings of EMNLP*.
- Oscar Méndez, Hiram Calvo, and Marco A. Moreno-Armendáriz. 2013. A reverse dictionary based on semantic analysis using wordnet. In *Proceedings of MICAI*.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Yuya Morinaga and Kazunori Yamaguchi. 2018. Improvement of reverse dictionary by tuning word vectors and category inference. In *Proceedings of ICIST*.
- Mohammad Taher Pilehvar. 2019. On the importance of distinguishing word meaning representations: A case study on reverse dictionary mapping. In *Proceedings of NAACL-HLT*.
- Fanchao Qi, Chenghao Yang, Zhiyuan Liu, Qiang Dong, Maosong Sun, and Zhendong Dong. 2019. Openhownet: An open sememe-based lexical knowledge base. *arXiv preprint arXiv:1901.09957*.
- Ryan Shaw, Anindya Datta, Debra E. VanderMeer, and Kaushik Dutta. 2013. Building a scalable database-driven reverse dictionary. *TKDE*, 25:528–540.
- Gerardo Sierra. 2000. The onomasiological dictionary: a gap in lexicography. In *Proceedings of the Ninth Euralex International Congress*.
- Sami Virpioja, Peter Smit, Stig Arne Grönroos, and Mikko Kurimo. 2013. Morfessor 2.0: Python implementation and extensions for morfessor baseline. *Aalto University Publication*.
- Lei Zhang, Fanchao Qi, Zhiyuan Liu, Yasheng Wang, Qun Liu, and Maosong Sun. 2020. Multi-channel reverse dictionary model. In *Proceedings of AAAI*.
- George Kingsley Zipf. 1949. Human behavior and the principle of least effort. *SERBIULA (sistema Librum 2.0)*.
- Michael Zock and Slaven Bilac. 2004. Word lookup on the basis of associations: from an idea to a roadmap. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries*.